

# VU Research Portal

## Storm Surge Reconstruction and Return Water Level Estimation in Southeast Asia for the 20th Century

Cid, Alba; Wahl, Thomas; Chambers, Don P.; Muis, Sanne

### ***published in***

Journal of Geophysical Research: Oceans  
2018

### ***DOI (link to publisher)***

[10.1002/2017JC013143](https://doi.org/10.1002/2017JC013143)

### ***document version***

Publisher's PDF, also known as Version of record

### ***document license***

Article 25fa Dutch Copyright Act

[Link to publication in VU Research Portal](#)

### ***citation for published version (APA)***

Cid, A., Wahl, T., Chambers, D. P., & Muis, S. (2018). Storm Surge Reconstruction and Return Water Level Estimation in Southeast Asia for the 20th Century. *Journal of Geophysical Research: Oceans*, 123(1), 437-451. <https://doi.org/10.1002/2017JC013143>

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### **E-mail address:**

[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)

## RESEARCH ARTICLE

10.1002/2017JC013143

## Key Points:

- A multiple linear regression model is used to reconstruct daily maxima surge levels at Southeast Asia from 1866 to 2012
- Return water levels are estimated using a GEV distribution and monthly maxima values
- The statistical model gives daily surge values of similar or higher quality than a global numerical model

## Supporting Information:

- Supporting Information S1
- Table S1

## Correspondence to:

A. Cid,  
cidcarrera@mail.usf.edu

## Citation:

Cid, A., Wahl, T., Chambers, D. P., & Muis, S. (2018). Storm surge reconstruction and return water level estimation in Southeast Asia for the 20th century. *Journal of Geophysical Research: Oceans*, 123, 437–451. <https://doi.org/10.1002/2017JC013143>

Received 29 MAY 2017

Accepted 16 DEC 2017

Accepted article online 27 DEC 2017

Published online 18 JAN 2018

## Storm Surge Reconstruction and Return Water Level Estimation in Southeast Asia for the 20th Century

Alba Cid<sup>1</sup> , Thomas Wahl<sup>2</sup> , Don P. Chambers<sup>1</sup> , and Sanne Muis<sup>3</sup> 

<sup>1</sup>College of Marine Science, University of South Florida, St. Petersburg, Florida, USA, <sup>2</sup>Department of Civil, Environmental and Construction Engineering & Sustainable Coastal Systems Cluster, University of Central Florida, Orlando, USA, <sup>3</sup>Institute for Environmental Studies (IVM), Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

**Abstract** We present a methodology to reconstruct the daily maximum storm surge levels, obtained from tide gauges, based on the surrounding atmospheric conditions from an atmospheric reanalysis (20th Century Reanalysis—20CR). Tide gauge records in Southeast Asia are relatively short, so this area is often underrepresented in studies based on long observational records, and there are just a few studies that have analyzed storm surge trends, variability or return water levels (RWLs) from numerical models in this area. Here we develop, calibrate, and validate a multivariate linear regression model that relates the storm surge with the principal components of the local atmospheric conditions. This allows us to reconstruct storm surges for the 147 year 20CR period (1866–2012) and therefore to calculate more robust RWLs from the entire simulated data set and subsets thereof. RWLs are obtained by fitting the monthly maxima values to the Generalized Extreme Value (GEV) distribution. We find an increase in the 50 year RWL from the second half of the 19th century to the present unrelated to mean sea level; this increase is less noticeable when comparing only recent periods. Therefore, further research is needed since there is evidence that atmospheric reanalyses can include spurious trends in the late 19th and early 20th. RWLs obtained from the statistical reconstruction are validated against the ones obtained from observations and from a numerical model. Agreements are generally higher when using surge levels from the statistical model, even before its calibration.

## 1. Introduction

Southeast Asia is highly vulnerable to, and frequently impacted by, extreme sea level events of different origin: typhoons and extra-tropical cyclones may both cause severe storm surges and rainfall with potentially devastating impacts to the economy and environment, and in many cases loss of human life. Typhoon Tiantu, for example, generated a 2 m storm surge along the Chinese Guangdong coast in 2013, resulting in losses of \$1.82 billion (Zhang & Sheng, 2015). Also in 2013, Typhoon Haiyan, an extremely intense tropical cyclone struck the Philippines, Vietnam, and nearby areas causing close to 6,000 fatalities and economic losses estimated at \$802 million (Mori et al., 2014). Tropical cyclone Nargis made landfall in Myanmar on May 2008; it was the worst natural disaster ever recorded affecting Myanmar, causing more than 138,000 fatalities and damage estimated at over \$10 billion. The storm surge peaked at over 5 m in the landfall area and inundated areas as far as 50 km inland (Fritz et al., 2009). Floods from these extreme events also have secondary effects in terms of groundwater contamination and saltwater intrusion (Cartwright et al., 2004; Pezeshki et al., 1990). Despite the large impact of extreme sea level events in the region, there are few studies that have analyzed the storm surge climate in this specific area, where large portions of the coastal population are among the most vulnerable to future changes in mean sea level and storminess.

Extreme sea levels (excluding tsunamis) emerge from the combination of astronomical tides, nontidal residuals and there is usually also a dynamic contribution of wind waves (which is not included in this study). The main contributor to nontidal residual extremes is the storm surge, even though the importance of the seasonal cycle and the steric component have been addressed in this work. Tides are deterministic and can be predicted, whereas information on storm surges need to be extracted from tide gauge (TG) records, if available, or modeled (numerically or statistically) by taking into account their forcing variables, such as wind stress and sea level pressure (SLP) gradients over the sea surface (e.g., Cid et al., 2016). In either case, high frequency (hourly or daily temporal resolution) and long storm surge records are required to analyze

trends and variability, and also to estimate return water levels (RWLs), which are key to coastal flood risk analyses and design.

Tide gauge records have been used to study sea level extremes, both at global and regional scales. At global scale, Menéndez and Woodworth (2010) investigated extreme sea level events and their spatial and temporal variability. They used monthly maxima values with the Generalized Extreme Value distribution (GEV) for studying the magnitude of extremes, and a Poisson process for analyzing the frequency. Merrifield et al. (2013) estimated annual maximum water levels for different sea level components (including the astronomical tide and high frequency nontidal residuals). Also at global scale, Marcos et al. (2015) investigated long-term variations in extreme sea level intensities and frequencies using a state-space model. They used long tide gauge records of at least 50 years from the GESLA data set (Menéndez & Woodworth, 2010). However, 90% of the tide gauges located in Southeast Asia have record lengths of less than 30 years. That is why the entire region is often underrepresented or not covered at all in such (near) global assessments based on long observational records. At regional scale there are only a few recent studies that have analyzed extreme sea levels from tide gauge records and have specifically obtained RWLs; those assessments were limited to the coast of China (Feng et al., 2015; Feng & Jiang, 2015; Feng & Tsimplis, 2014).

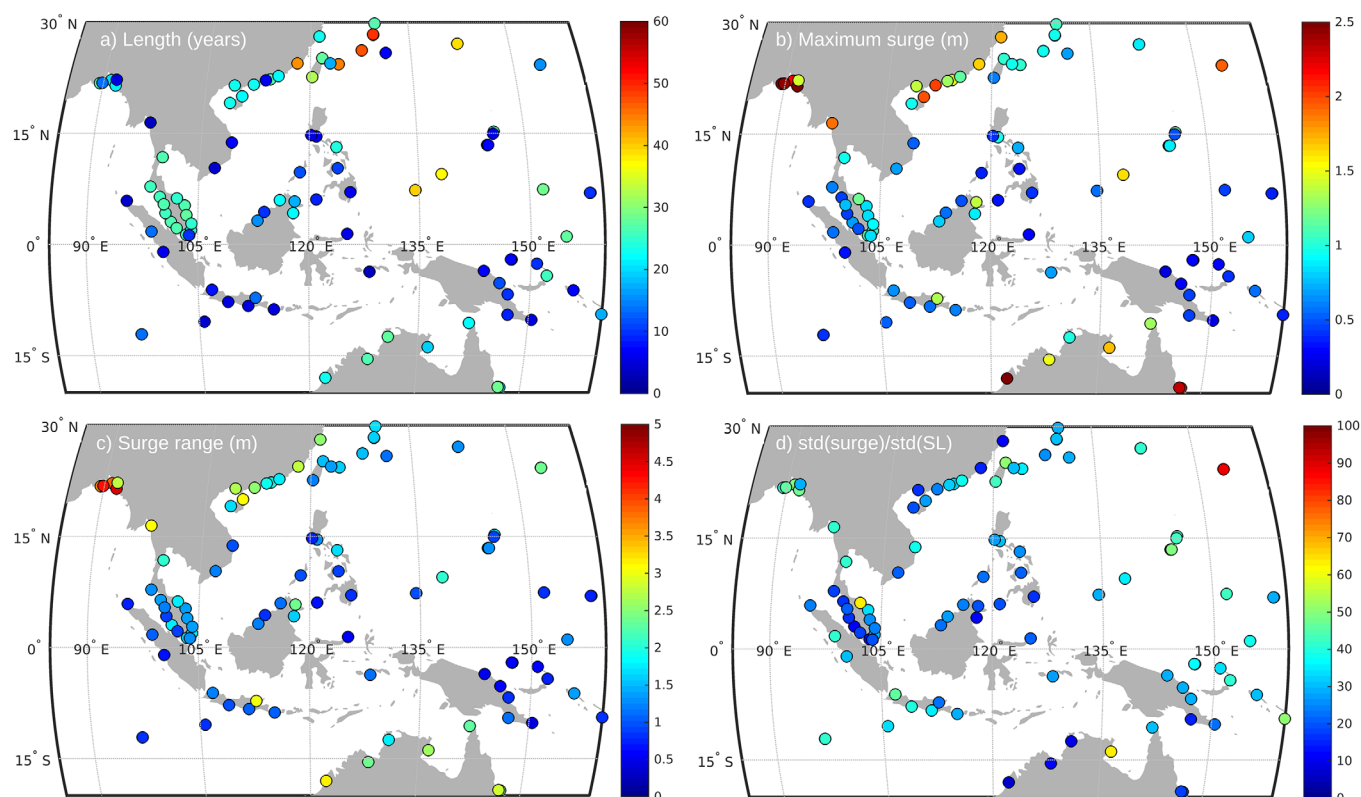
One way to overcome the absence of long tide gauge records is to use numerical models to simulate the storm surge component. At global scale, Muis et al. (2016) used the Delft3D Flexible Mesh software (Kernkamp et al., 2011) to simulate storm surge data from 1979 to 2014 (Global Tide and Surge Reanalysis—GTSR) at a resolution of approximately 5 km close to the coast. Based on the 36 year reanalysis, they obtained RWLs by fitting the annual maxima values to a Gumbel distribution. At regional scale, we are only aware of the work by Zhang and Sheng (2015), who used the Princeton Ocean Model (POM) (Mellor, 2004) for obtaining hourly storm surge values. From the latter, they derived annual maxima and determined 50 year RWLs along the Northwest Pacific coasts of China, Japan, and Russia by fitting a Gumbel distribution. Other papers have also analyzed modeled storm surge data in the area, but only at local scale (e.g., Guo et al., 2009; Wang et al., 2005; Zhang et al., 2007). A common drawback of numerical models is that they need accurate high-resolution bathymetric data to resolve small-scale near-shore processes. Such data is often not available.

A third approach, complementary to the use of (long) tide gauge records or numerical models, is to develop and apply statistical models to reconstruct the storm surge component. In this work, we use a multivariate regression model to reproduce the storm surge signal measured by tide gauges across Southeast Asia. Statistical techniques have been widely used to reconstruct wave-related parameters (Camus et al., 2016; Rueda et al., 2016; Wang et al., 2012), but have also been applied to reproduce the storm surge. For example, Dangendorf et al. (2014) used a statistical-empirical formulation proposed by Müller-Navarra and Giese (1999) to analyze the long-term behavior of storm surges in the North Sea and to assess the validity of reanalysis atmospheric data to reconstruct storm surges in the late 19th and early 20th centuries. Wahl and Chambers (2016) used linear regression models to explain the observed multidecadal storm surge variability along the U.S. coast by means of traditional and tailored climate indices. Cid et al. (2016) extended the temporal coverage of the global DAC database (Dynamic Atmospheric Correction from AVISO) using tailored indices obtained from atmospheric fields from the 20th Century Reanalysis (Compo et al., 2011). We are not aware of local or regional studies where statistical models have been used to analyze extreme sea levels in Southeast Asia. Hence, our study has the following three main objectives: (1) to apply a similar approach as used in Cid et al. (2016) to develop, calibrate, and validate a statistical model to simulate daily and monthly maxima values of storm surge at tide gauge locations in Southeast Asia; (2) to use the model to simulate storm surges for the 147 year period from 1866 to 2012; and (3) to calculate more robust return water levels from the entire simulated data set, and subsets thereof to assess potential changes between different periods.

The remainder of the paper is organized as follows: the TGs and atmospheric data sets are introduced in section 2. The methodology for the model development, validation and bias correction, as well as the extreme value analysis are described in section 3. Results are presented and discussed in section 4 and the main conclusions are summarized in section 5.

## 2. Data

We use hourly sea level records from the GESLA 2 database (Woodworth et al., 2017, <http://www.gesla.org>). All available tide gauges within the area enclosed by 85°E and 160°E and between 30°N and 20°S (Figure 1)



**Figure 1.** Location and description of the available GESLA2 tide gauge records for Southeast Asia (133 TGs). (a) Length (years) of the tide gauge records. (b) Maximum storm surge value (m) recorded at each tide gauge station. (c) Storm surge range (m) expressed as the difference between the maximum and minimum values. (d) Storm surge standard deviation over the total sea level standard deviation (%).

were downloaded (158 TG records) and preprocessed to remove isolated spikes, and doubtful or wrong values. After the data quality control, only those TGs with at least 2 years of valid data were retained (one for calibrating the model and the other one for the validation of the results), making 133 in total (Figure 1). The longest records come from stations located along the coast of Japan and Xiamen in China, with lengths over 40 years. Approximately 50% of the records cover less than 20 years.

We obtain the astronomical tide from a year-by-year classical tidal harmonic analysis using the *T-Tide* package (Pawlowicz et al., 2002), accounting for 67 constituents. Years with less than 75% of completeness were excluded from the analysis. Nontidal residual time series are obtained by subtracting the astronomical tide from the observed sea level. Mean sea level variability is removed by subtracting the annual moving median at each station.

It is worth mentioning that the nontidal residual potentially includes a steric component, which is not directly related to storm surge. A test removing the monthly mean steric signal was carried out with no improvement in the results (see supporting information for more details). Therefore, in this work, we assume that the nontidal residual obtained after removing both the astronomical tide and mean sea level variability is equivalent to the storm surge signal. Relevant storm surge characteristics in the study area are summarized in Figures 1b–1d. Areas frequently affected by tropical cyclones have the largest surge magnitudes, such as the head of the Bay of Bengal, the southernmost coast of China and the Northeast and Northwest coast of Australia, where the storm surges reach magnitudes of around 2 m (Figure 1b). The storm surge range, expressed as the difference between the maximum and minimum values recorded at each station (Figure 1c), shows that natural variability is mostly lower than 2 m, except for stations located in the Bay of Bengal, China, and Australia. Figure 1d shows the storm surge variation relative to the total sea level variation, expressed as the standard deviation of the storm surge over the standard deviation of the total sea level; on average, the nontidal residual accounts for around 30% of the total sea level variability.

The 20th Century Reanalysis version 2c ensemble mean (20CRv2c) (Compo et al., 2011) was used to select the atmospheric variables of interest: the pressure at mean sea level (SLP) and the meridional and zonal wind components at 10 m (Uwind and Vwind, respectively). SLP has a temporal resolution of 6 and spatial resolution of  $2^\circ$  on a regular grid; wind fields have 3 hourly resolution and are available on a Gaussian grid with a spatial resolution of  $\sim 1.875^\circ$ . The wind components were interpolated onto the  $2^\circ$  SLP grid, as well as interpolated to a 6 hourly temporal resolution.

Multiple tests were carried out using different parameters and (sub-)sets (see section 3.1 and supporting information for more details); using the ones described above led to the best results. The 20CRv2c spans the period 1851–2014, but a pressure bias has been reported from 1851 to 1865 ([https://www.esrl.noaa.gov/psd/data/gridded/20thC\\_ReanV2c/opportunities.html](https://www.esrl.noaa.gov/psd/data/gridded/20thC_ReanV2c/opportunities.html)) and the database was experimental for the 2013–2014 period at the time of analysis. Therefore, in this work, we consider the 20CRv2c atmospheric variables for the 147 year period from 1866 to 2012.

### 3. Methods

As previously mentioned, the objective of this work is to reconstruct the storm surge signal as measured by tide gauges for the 20CRv2c period, with the intention of estimating reliable RWLs for Southeast Asia for different time periods. This is done by obtaining the statistical relationship, through a multivariate linear regression model, between the daily maximum surge levels (predictand) and the surrounding atmospheric conditions (predictor).

#### 3.1. Predictor and Predictand Definition

Different tests were carried out using different predictors (both in terms of the area covered by the predictor and the variables taken into account) and also different predictands (taking/not taking into account the seasonal cycle, the inverse barometer effect, etc.); see the supporting information section for a full description.

Bests results were found when using a local predictor and removing the seasonal cycle from both the predictor and the predictand. The seasonal cycle was removed by subtracting climatological months of nontidal residual (predictand) and climatological months of mean sea level pressure (predictor). The local predictor is based on the 10 m meridional and zonal winds, the sea level pressure fields, and zonal and meridional gradients of the SLP (GRDslp). Only those grid points surrounding the tide gauge location were used, specifically, an area of  $4^\circ \times 4^\circ$  ( $3 \times 3$  grid cells) centered at the tide gauge location. Regarding the predictor, we select the daily minimum pressure and the winds and gradients associated to that instant; and regarding the predictand, daily maximum values were only calculated for complete days (i.e., 24 values available).

Once the predictor area and variables are defined, the next step consists in performing a Principal Component Analysis (PCA) to the predictor, made of nine grid cells for each of the four variables (36 components in total), and keep only the Principal Components (PCs) that explain 95% of the variance to reduce the dimensionality of the model. Since the annual moving median was removed from the nontidal residual, the same process is applied to the obtained PCs.

To summarize, the predictand in our model consists of daily maximum values of surge levels without the seasonal cycle. The selected predictor is comprised of the PCs obtained from the daily minimum SLP (without seasonal cycle) and associated GRDslp, Uwind, and Vwind of the  $3 \times 3$  grid cells centered at the tide gauge location. By removing the seasonal cycle before fitting the model we increase its sensitivity and performance, but the seasonal cycle is added back to the reconstructed surge prior to the extreme analysis and bias correction.

#### 3.2. Model Development

The statistical method to reconstruct the surge levels is explained in detail in Cid et al. (2016), which was in turn based on a similar approach used by Camus et al. (2014) to downscale multivariate wave climate. It consists of fitting a multivariate regression model between daily maximum surge levels (predictand) and the PCs of the daily SLP, GRDslp, and wind (predictor)

$$\text{surge}(x_i, t) = a_i + b_{1,i} \times PC_{(1)}(x_i, t) + b_{2,i} \times PC_{(2)}(x_i, t) + \dots + b_{n,i} \times PC_{(n)}(x_i, t) \quad (1)$$

where  $n$  is the number of PCs that achieved a statistical improvement of the results and  $a_i, b_{1,i}, \dots, b_{n,i}$  are the coefficients obtained from the regression model.

The model is fitted in a forward procedure: the first predictor PC is obtained from the best fit (smallest sum of squared errors) among each of the components separately ( $PC_{(1)}$  in equation (1)). The second predictor PC ( $PC_{(2)}$  in equation (1)) is chosen from the rest of the PCs so that it gives the best fit with two predictors, the best predictor selected in the previous model plus one of the remaining potential predictors. The cycle continues until a more complex model does not produce a significant improvement (F-test at the 5% level of significance) in the multivariate regression fit. Therefore, only those PCs that produce a significant improvement in the predicted surge are taken into account in the statistical model.

This methodology allows us to estimate surge levels as a linear combination of the most important PCs. As can be seen from equation (1), surge levels at any given location ( $x_i$ ) can be estimated from a specific number of PCs, which varies spatially.

Finally, when the coefficients of the statistical model are obtained, the 147 years of atmospheric data from 20CRv2c are used to reconstruct the daily maxima surge levels, giving a continuous storm surge time series, without gaps, from 1866 to 2012 at each tide gauge location.

### 3.3. Model Validation and Bias Correction

A k-fold cross-validation process is used, i.e., the regression model is fitted for all years except one. The latter is used to validate the surge reconstruction, and the process is repeated for the total length of the tide gauge record. Once the model is validated, the coefficients of the linear regression model are obtained by fitting the complete surge record at each location using the overlapping period from the 20CRv2c predictors.

Both hydrodynamic and statistical models often exhibit a bias when compared to observed data (e.g., Arns et al., 2015; Dangendorf et al., 2014). Here we remove this bias from monthly maxima values (which are used for the extreme value analysis, see section 3.4) by applying quantile-mapping (e.g., Boé et al., 2007). The latter calibrates the Cumulative Distribution Function (CDF) of the simulation results; i.e., the empirical quantiles are obtained for both the observed and simulated surge, and the differences are added to the simulated distribution.

### 3.4. Extreme Value Analysis

After we have tested the ability of the statistical model to reconstruct the observed monthly maxima storm surges we fit the monthly maxima values to a GEV distribution (see equation (2)).

By choosing monthly maxima to fit the extreme model, we have enough observation values (also at sites with shorter records) to estimate the 50 year RWLs and can compare results from observations with the ones derived from the reconstruction. An additional validation was done by comparing with the 50 year RWLs obtained from GTSR monthly maxima values

$$G(x) = \begin{cases} \exp\left(-\left[1 + \xi\left(\frac{x-\mu}{\psi}\right)^{\frac{1}{\xi}}\right]\right) & \xi \neq 0 \\ \exp\left(-\exp\left[-\left(\frac{x-\mu}{\psi}\right)\right]\right) & \xi = 0 \end{cases} \quad (2)$$

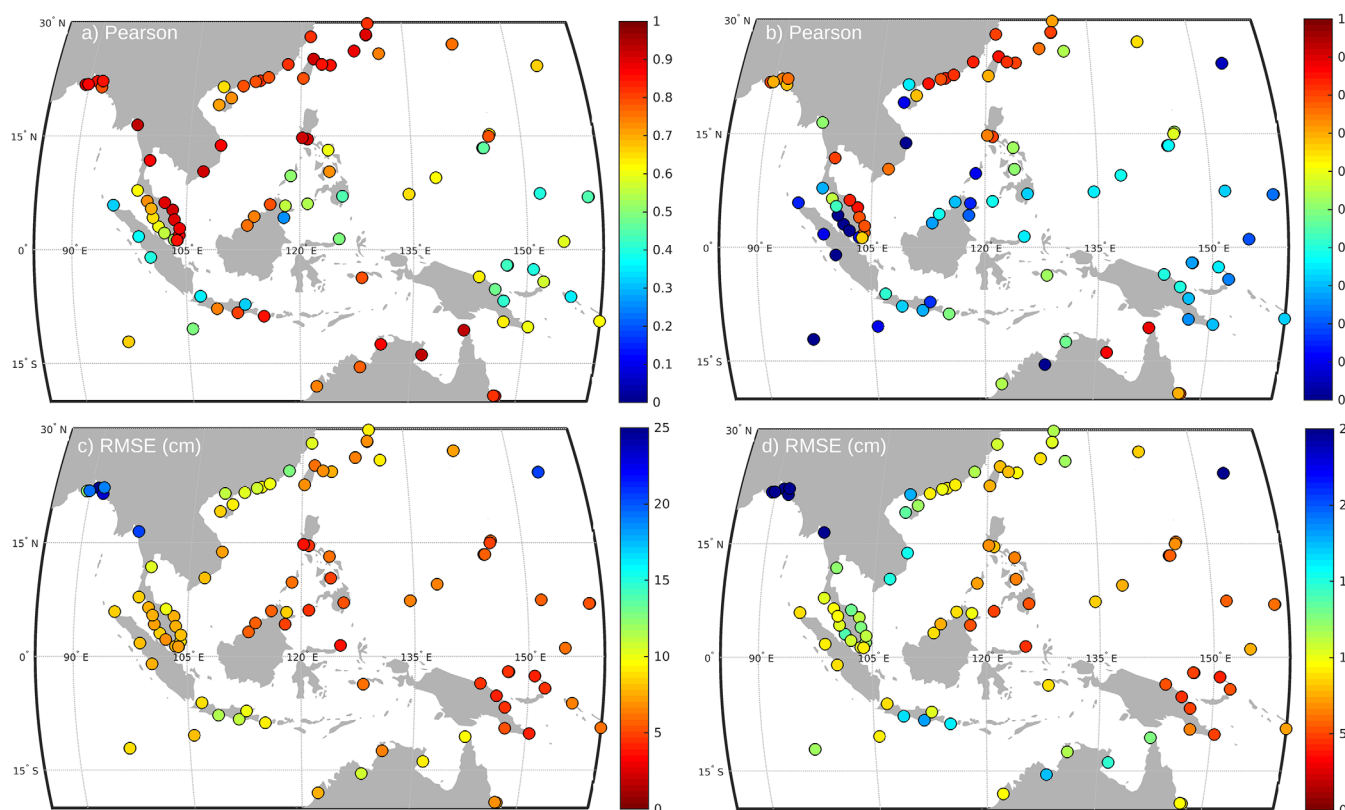
where  $-\infty < \mu < \infty$  is the location parameter,  $\psi > 0$  is the scale parameter, and  $-\infty < \xi < \infty$  is the shape parameter. Parameters are estimated by the maximum likelihood method.

## 4. Results

### 4.1. Validation of Daily Maxima

Validation results are shown in Figures 2a and 2c in terms of the Pearson correlation coefficient and the root mean squared error (RMSE). The comparison between observations and the results from a numerical model (GTSR) is also depicted (Figures 2b and 2d). The first thing to notice is that the statistical model gives

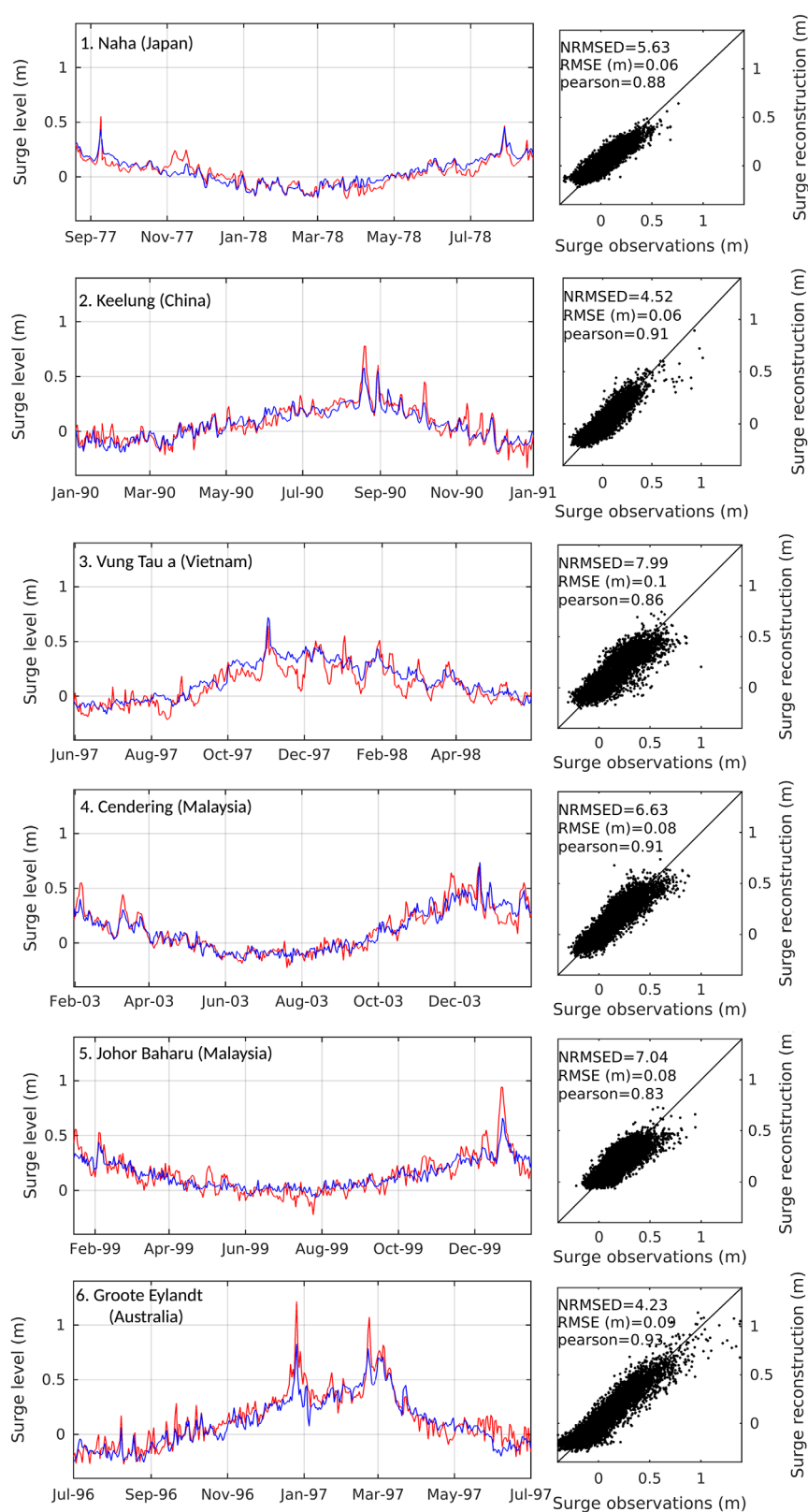




**Figure 2.** Validation of the daily maxima values in terms of (top) the Pearson coefficient and (bottom) the RMSE (cm). Agreement between observations and the statistical reconstruction after adding back the (left) seasonal cycle and (right) between observations and GTSR.

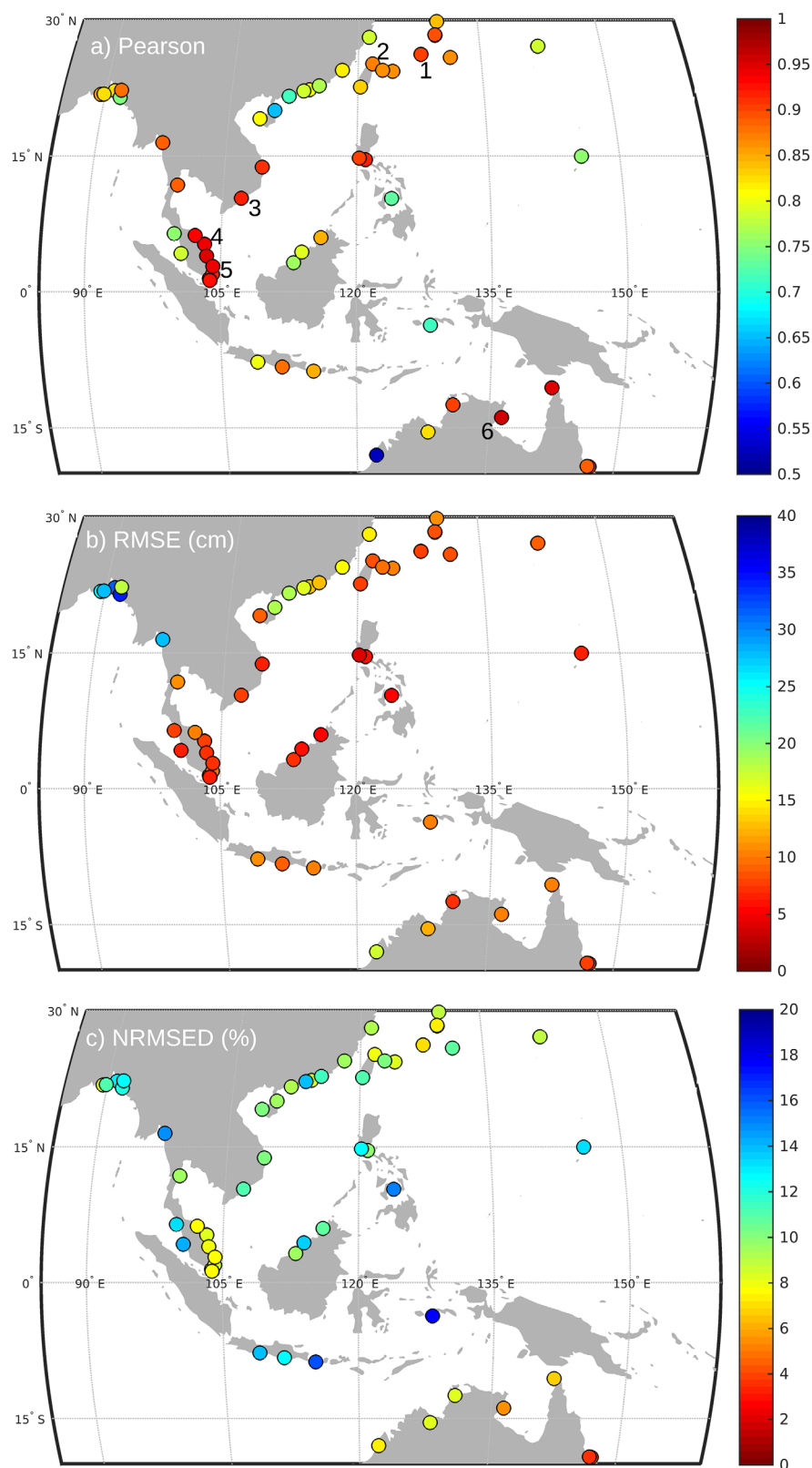
a higher agreement with observations than the numerical model, and that the areas of common good/bad performance agree between models. This shows that the relationship between the nontidal residual and the atmospheric signal is less strong at those locations. The model performance is more dependent on the latitude (generally, with smaller performance at lower latitudes) than for example on the record length or on the storm surge range. The Pearson coefficient is above 0.7 at 50% of tide gauge locations (Figure 2a). Lower model performance is found at those tide gauges located in semienclosed areas (e.g., Strait of Malacca) or areas with an intricate coastal morphology (e.g., West Sumatra coast). In those areas, the local and baroclinic effects become more important and the  $2^\circ$  resolution of the atmospheric variables is not a good descriptor of the local atmospheric conditions. Results are also less accurate for small islands, where, for example, the sea level variation due to eddies or planetary waves can be gathered in the nontidal residual, which could explain that the variability is higher in the observations than in the reconstruction, especially at low latitudes. The RMSE is close to or under 10 cm at all tide gauge locations except for those situated at the head of the Bay of Bengal, where errors range between 15 and 25 cm (Figure 2c). This is an area where generally greater surge magnitudes are found, as can also be seen from Figure 1b. The statistical model performs better along the head of the Bay of Bengal, the coasts of East Malaysia, Thailand, Vietnam, China, and Japan.

Figure 3 depicts a time series comparison between the observed and reconstructed daily maximum nontidal residuals for 1 year periods at six locations scattered throughout the investigation area (see location in Figure 4a). Scatter plots are also shown for the entire periods. Both the time series and scatter plots indicate good performance of the statistical model, with correlation coefficients varying between 0.83 at Johor Baharu (Malaysia) and 0.93 at Groote Eylandt (Australia), and with maximum relative errors (RMSE over the surge range) between 4% at Groote Eylandt and 8% at Vung Tau (Vietnam). It can be seen from the upper tail of the scatter plots that extreme values are often underestimated. This is mainly because the linear regression model is fitted at a daily scale, using the smallest sum of squared errors as the indicator of the best agreement, which leads to a good fit for the daily values but less accurate results for the most extreme events.



**Figure 3.** Daily maximum surge (m). Time series (1 year period) comparison (before bias correction) and scatter plots (full period) at six different locations. Red line represents the observed surge obtained from tide gauges and the blue line represents the statistical reconstruction. See TG locations in Figure 4.





**Figure 4.** Validation of the monthly maxima values (after bias correction) for those tide gauges where the validation of the daily maxima gives a Pearson coefficient  $>0.7$  and an NRMSED  $<15\%$ . (a) Pearson coefficient. (b) RMSE (cm). (c) NRMSED (%). Numbers 1–6 locate the TGs depicted in Figures 3, 5, and 8.

#### 4.2. Validation and Bias Correction of Monthly Maxima

We have seen that the highest surge events are often underestimated, therefore a correction should be carried out before analyzing extreme events. Here we apply the quantile-mapping bias correction described in section 3.3.

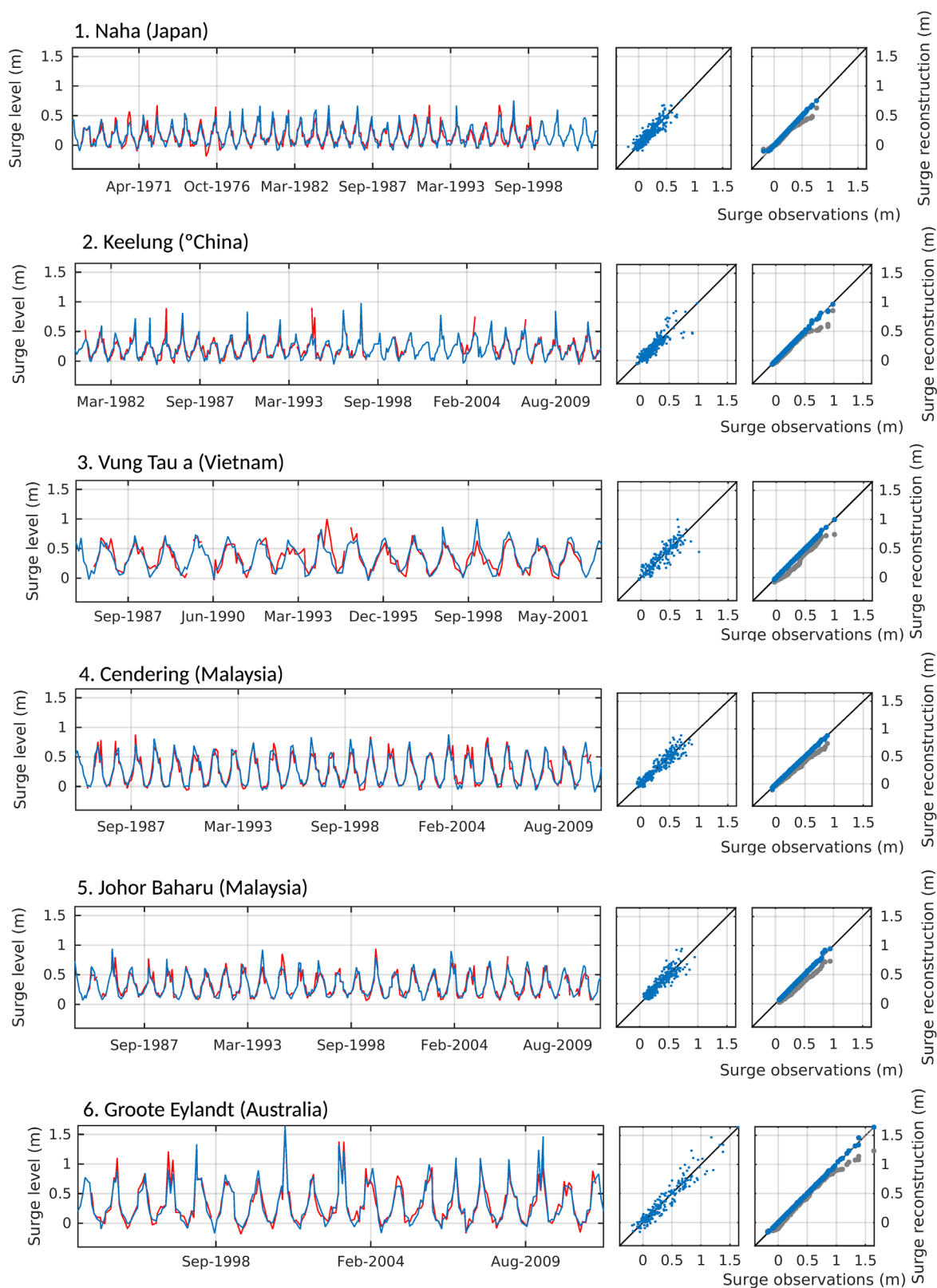
Tide gauges, where the validation of the daily maxima values gives Pearson coefficients lower than 0.7 or NRMSED greater than 15%, are disregarded at this stage and not subjected to the extreme value analysis described in the following section 4.3. For the tide gauges where the validation meets this quality standard (65 TGs in total), monthly maxima values are obtained (if the monthly completeness is  $>60\%$ ; Menéndez & Woodworth, 2010) and bias corrected. The agreement between the observed and the reconstructed monthly maxima surges after the bias correction is shown in Figure 4. TGs at the east coast of Malaysia and in the Gulf of Carpentaria show the highest agreement, reaching Pearson coefficients over 0.9 (see Figure 4a) and NRMSED below 9% (see Figure 4c). Although the head of the Bay of Bengal presents high correlation coefficients, the differences between the observed and reconstructed monthly maxima can reach more than 25 cm (see Figure 4b), but this only equals errors around 15% with respect to the surge variability in that area.

Figure 5 shows the time series comparison, scatter plots, and quantile-quantile plots (q-q plots) between the observed and the bias corrected monthly maxima for the six locations highlighted in Figure 3. The q-q plots also show results before bias corrections (grey dots). The intra-annual variability is captured by the statistical model and the extreme values are generally well reproduced. Cendering and Grote Eyland both have Pearson coefficients of 0.94 and NRMSED of 8% and 6%, respectively. Johor Baharu, Keelung, and Naha have slightly lower correlation coefficients (0.91, 0.87, 0.9, respectively) and slightly higher relative errors (9%, 8%, 7%).

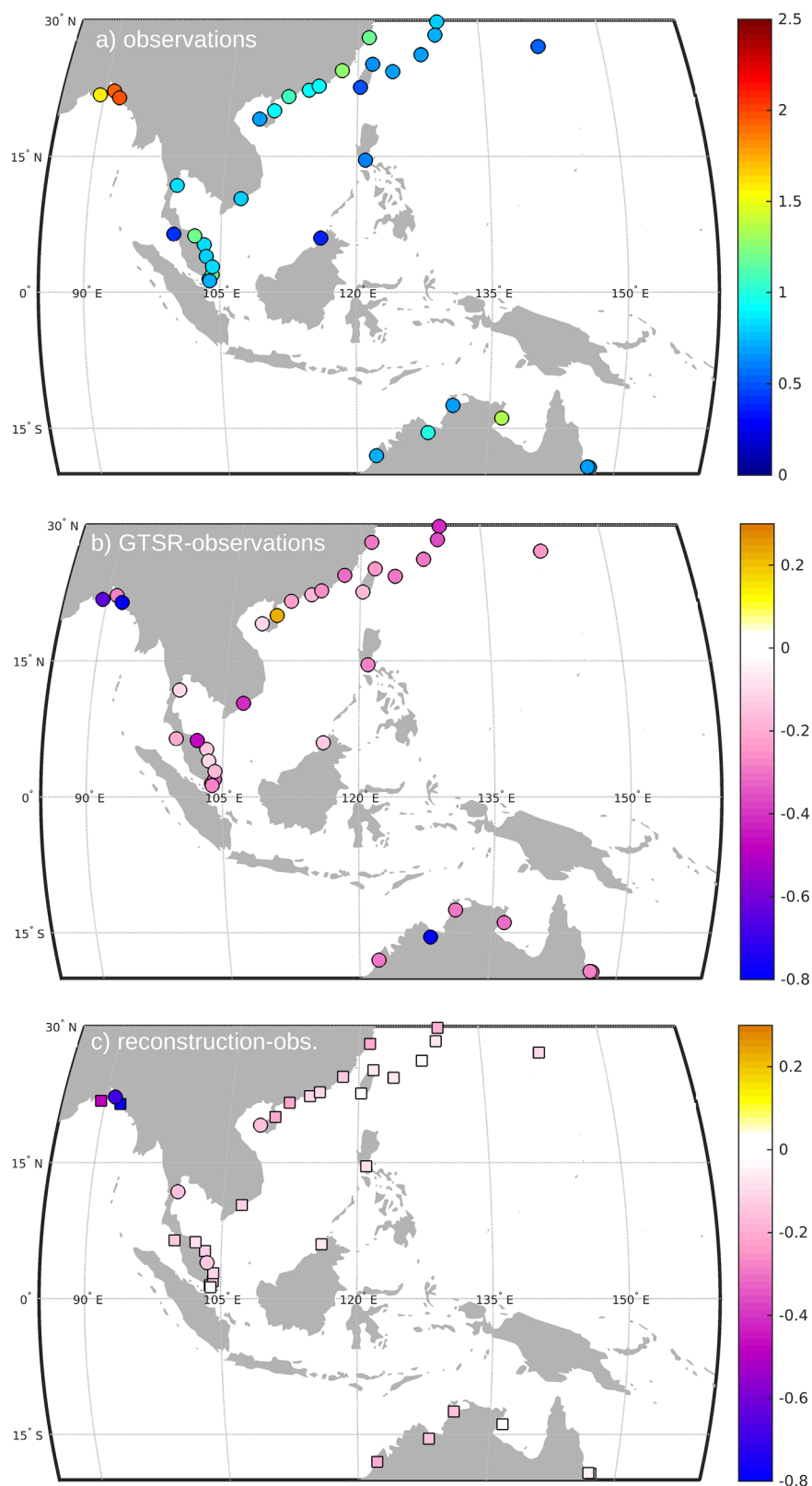
#### 4.3. Extreme Value Analysis

In order to further validate our results, we calculate 50 year RWLs (using the GEV and monthly maxima) at stations with records longer than 15 years (Figure 6a), noting that extrapolation should be limited to return periods not longer than four times the observational record (Pugh & Woodworth, 2014). We compare those estimates from observations to the ones obtained from the reconstruction (Figure 6c) and also to the ones we obtained using surge levels from GTSR (Muis et al., 2016; Figure 6b). RWLs are generally underestimated by the two models. Differences between the 50 year RWLs from observations and our model results range from 2 to 80 cm and are 13 cm on average (before the bias correction). The largest differences are observed again in the Bay of Bengal where storm surges are highest. Differences in the 50 year RWLs from observations and GTSR provide a similar spatial pattern but are generally higher; squares in Figure 6c highlight sites where our model (before bias correction) leads to more accurate results than GTSR. In terms of both the spatial pattern and magnitude of RWLs, our results are similar to the ones obtained by Feng and Tsimplis (2014) for the coast of China and Zhang and Sheng (2015) for the northwestern Pacific coast. Differences may result from different models used to simulate water levels and different approaches used for the extreme value analysis.

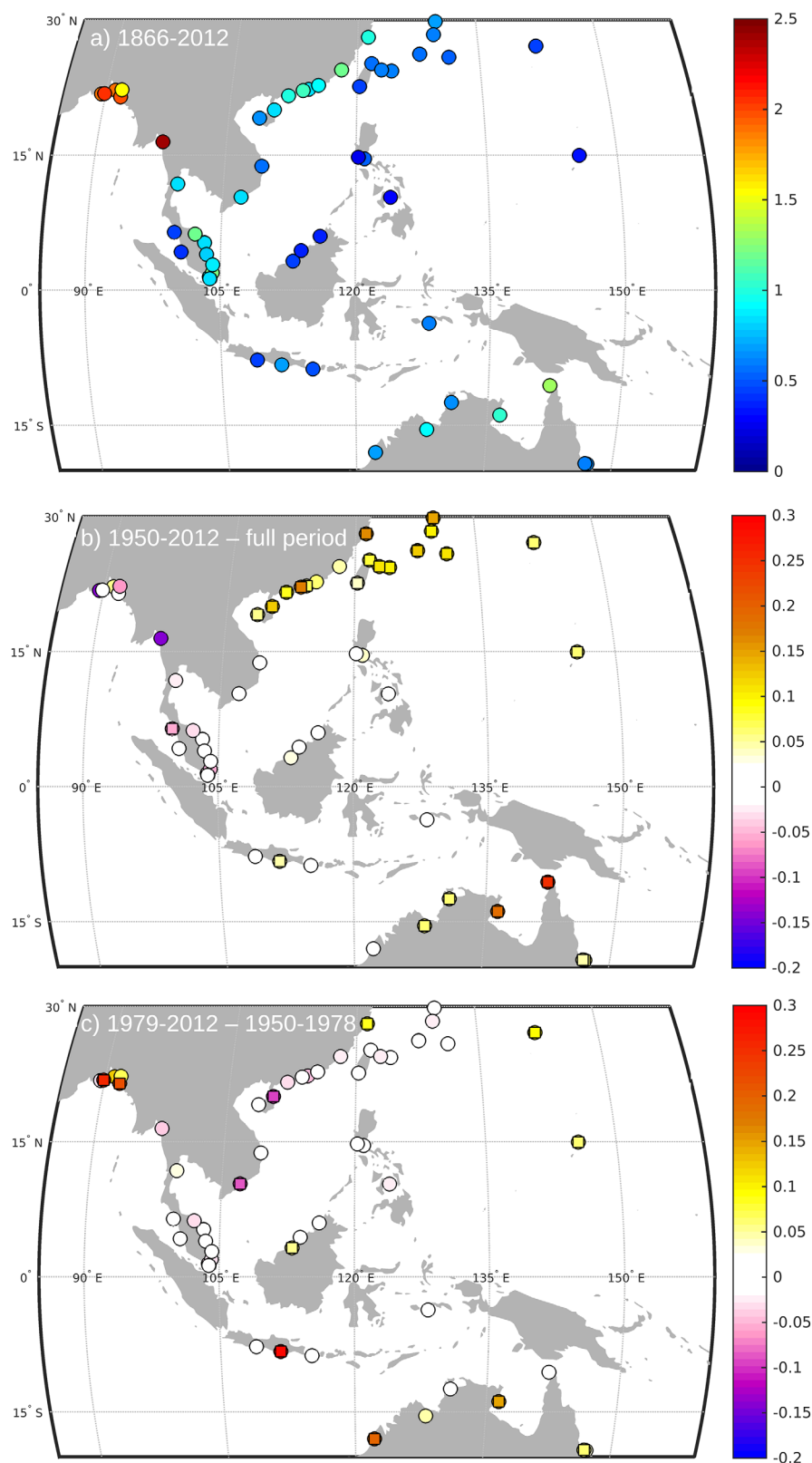
After this additional validation exercise, we reconstruct surge water levels (at tide gauges with Pearson coefficients  $>0.7$  and NRMSED  $<15\%$ ) for the entire period from 1866 to 2012 using the 20CRv2c atmospheric predictors, add back the seasonal cycle, compute monthly maxima values and apply the quantile-mapping correction. The 50 year RWLs obtained from monthly maxima over the entire period are shown in Figure 7a. We also calculated RWLs for the 1950–2012 period and for two subsets of approximately 30 years, the 1950–1978 period and 1979–2012 period (presatellite and postsatellite era, respectively). Differences in the 50 year RWLs between the entire simulated data set and subsets are depicted in Figures 7b and 7c, respectively. Squares represent the locations where the RWL differences between the subsets and the full period are greater than the confidence interval for the full period. An increase in the 50 year RWLs can be seen from the second half of the 19th century to the present (Figure 7b) along the East coast of China and Japan, and North Australia, with increases close to 25 cm in the Gulf of Carpentaria. When comparing more recent periods (Figure 7c) we see that the increase is spatially less consistent and that most stations do not present differences in the 50 year RWL; so based on the analyzed TG locations, it is not clear whether the 20CRv2c has a lower quality in the presatellite era (1950–1978) for the Southeast Asia region. There are small differences between these two recent periods along the coasts of China, Japan, and Australia; while big differences (over 20 cm) are found at the Gulf of Bengal. Stations along Vietnam, Malaysia, and Indonesia present the smallest variations, with differences below 5 cm. There are only a few stations that show a decrease in the 50 year RWL in the last 60 years, this reduction is close to 10 cm only at Vung Tau (Vietnam) and Haikou



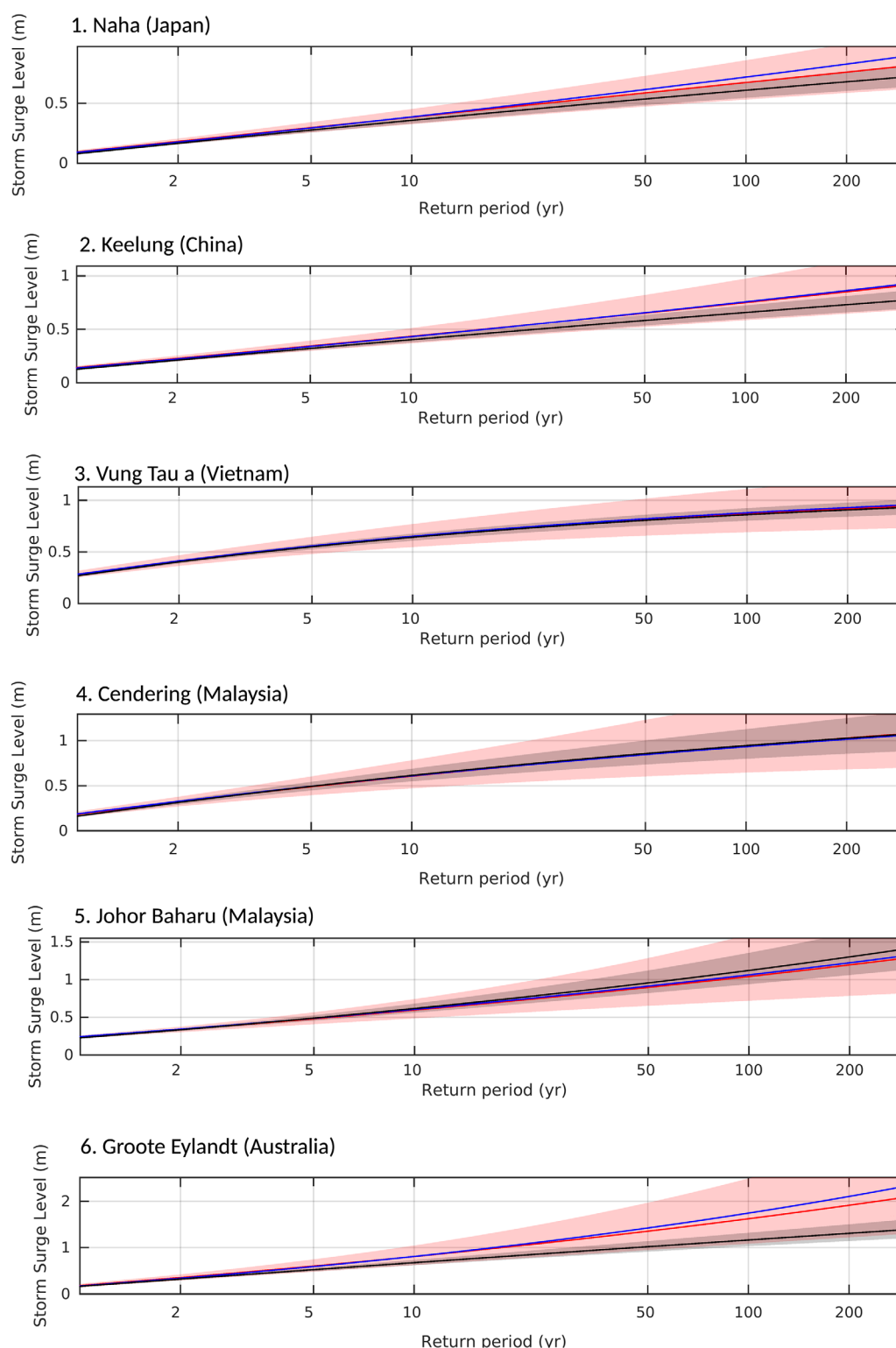
**Figure 5.** Validation of the monthly maxima values (after correction). Time series comparison between observed monthly maxima (red line) and reconstructed monthly maxima (blue line), scatterplot and q-q plot. Grey dots on the q-q plots represent the quantiles before the bias correction.



**Figure 6.** Monthly maxima values fitted to a GEV distribution for those tide gauges longer than 15 years. (a) 50 year RWL (m) estimated from observations. (b) Difference in the 50 year RWL between GTSR and the observations (for the observations period). (c) Difference in the 50 year RWL between the reconstruction before the bias correction and the observations (for the observations period). Squares represent the locations where the reconstruction before correction is more accurate than GTSR.



**Figure 7.** Monthly maxima values fitted to a GEV distribution. (a) 50 year RWL (m) estimated from the reconstruction for the full period (1866–2012). (b) Difference between the 50 year RWL for the 1950–2012 period and (a). (c) Difference between the 50 year RWL for the 1979–2012 period and the 1920–1978 period. Squares represent the locations where the RWL differences between the subset and the full period are greater than the confidence interval for the full period.



**Figure 8.** RWLs (m) estimated from the fit of the monthly maxima values to a GEV distribution. Red line represents the estimation of the RWLs from observations (TGs with at least 15 years of data) and the red shaded area represents the confidence intervals. Blue line represents the RWLs estimated from the reconstruction (for the observation's period). The black line corresponds to the RWLs estimated from the reconstruction taken into account the complete 1866–2012 period, the grey shaded area represents the confidence intervals.



(China). Differences in the 50 year RWL are significant (i.e., difference is bigger than the confidence intervals for the full period) at 32 TG locations when comparing the baseline period and the 1950–2012, and at 17 locations when comparing the 1950–1978 and the 1979–2012 periods. However, those results have to be interpreted with great caution since there is evidence that reanalysis products, such as 20CRv2c, can include spurious trends in the early parts where little observational data was available (Dangendorf et al., 2014; Krueger et al., 2013).

An obvious effect of having longer records available for the extreme value analysis is a reduction of uncertainties in estimated RWLs. This is highlighted in Figure 8 where GEV distributions and confidence levels are shown for observations (red) and for simulation results for 1866–2012 (black and grey). At some sites, such as Groote Eylandt, the changes in RWLs over time are evident again, in addition to the narrowing confidence levels. As mentioned above it has not been tested (due to the lack of centennial tide gauge records) how reliable our simulation results are for the late 19th and early 20th century; but even calculating RWLs for the last 50 or 60 years (where it has been shown in other regions that reanalysis products provide reasonable results), instead of much shorter observational periods, leads to a significant reduction in uncertainties. Finally, Figure 8 again shows the good agreement between observations (red) and simulations (blue) when the same time period is considered.

## 5. Conclusions

This study presents a methodology to reconstruct the daily maximum storm surge levels in Southeast Asia, obtained from tide gauges, based on the surrounding atmospheric conditions from an atmospheric reanalysis (20CRv2c). This is done by means of a multivariate linear regression model that relates the storm surge with the PCs of the local atmospheric conditions.

The model validation has shown overall good results, but also revealed weaknesses around small islands with complicated coastal morphology (e.g., West Sumatra coast), and in semiencloded basins (e.g., Malacca Strait) where baroclinic and other effects (not covered by our model) become more important. Comparing results from our statistical model with those from a global hydrodynamic storm surge model shows that the former outperforms the latter at most locations in Southeast Asia. Numerical models, such as the one used to produce the GTSR data set, heavily rely on accurate high-resolution bathymetric data, which is often not available. It has been shown here that in those cases statistical models can be a very useful addition to more sophisticated process based models. Both types of models rely on the accuracy of the atmospheric forcing, where tropical cyclones are generally poorly resolved.

Process based models are also computationally very expensive, whereas statistical models can produce long records very fast for locations where observational records are available. In the present study, we have used atmospheric data from 20CRv2c to reconstruct storm surges over a 147 year period from 1866 to 2012. Using extreme value analysis, it has been shown how such long data sets can be used to detect changes in extreme sea levels (while being aware of potential shortcomings in the atmospheric forcing in the late 19th and early 20th century) and help determining more robust RWLs that are highly important for coastal flood risk management and adaptation. In future research, we will use a similar approach to better validate reanalysis atmospheric conditions at centennial time scales based on long tide gauge records. We will explore the applicability of the presented methodology to other regions where limited observational data availability has hampered a better understanding of extreme sea levels and RWLs and potential changes related to climate variability and change. For coastal flooding risk, it is also important to have continuous information around the coastline, not only at TG locations, so we will combine the statistical and hydrodynamic (e.g., GTSR) approaches to apply the statistical model at a global scale and reduce the uncertainty in the RWLs compared to the reanalysis.

## References

- Arns, A., Wahl, T., Haigh, I. D., & Jensen, J. (2015). Determining return water levels at ungauged coastal sites: A case study for northern Germany. *Ocean Dynamics*, 65(4), 539–554. <https://doi.org/10.1007/s10236-015-0814-1>
- Boé, J., Terray, L., Habets, F., & Martin, E. (2007). Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies. *International Journal of Climatology*, 27(12), 1643–1655. <https://doi.org/10.1002/joc.1602>
- Camus, P., Méndez, F. J., Losada, I. J., Menéndez, M., Espejo, A., Pérez, J., . . . Guanche, Y. (2014). A method for finding the optimal predictor indices for local wave climate conditions. *Ocean Dynamics*, 64(7), 1025–1038. <https://doi.org/10.1007/s10236-014-0737-2>

## Acknowledgments

We would like to thank Ivan Haigh for providing the steric sea levels derived from the Ishii et al. (2006) data set that were used here for testing purposes. We also thank the three anonymous reviewers who helped to improve an earlier version of the manuscript. Monthly maxima surge values from GTSR, for the selected tide gauge locations, are available at <https://doi.org/10.5281/zenodo.584139>. The tide gauge data used in this work are freely available through the GESLA2 website (<http://www.gesla.org>). Support for the Twentieth Century Reanalysis Project version 2c dataset is provided by the U.S. Department of Energy, Office of Science Biological and Environmental Research (BER) and by the National Oceanic and Atmospheric Administration Climate Program Office. 20th Century Reanalysis V2c data were provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, through their web site <http://www.esrl.noaa.gov/psd/>.

- Camus, P., Rueda, A., Méndez, F. J., & Losada, I. J. (2016). An atmospheric-to-marine synoptic classification for statistical downscaling marine climate. *Ocean Dynamics*, 66(12), 1589–1601. <https://doi.org/10.1007/s10236-016-1004-5>
- Cartwright, N., Li, L., & Nielsen, P. (2004). Response of the salt-freshwater interface in a coastal aquifer to a wave-induced groundwater pulse: Field observations and modelling. *Advances in Water Resources*, 27(3), 297–303. <https://doi.org/10.1016/j.advwatres.2003.12.005>
- Cid, A., Camus, P., Castanedo, S., Méndez, F. J., & Medina, R. (2016). Global reconstructed daily surge levels from the 20th Century Reanalysis (1871–2010). *Global and Planetary Change*, 148, 9–21. <https://doi.org/10.1016/j.gloplacha.2016.11.006>
- Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., . . . Worley, S. J. (2011). The Twentieth Century Reanalysis Project. *Quarterly Journal of the Royal Meteorological Society*, 137(January), 1–28. <https://doi.org/10.1002/qj.776>
- Dangendorf, S., Müller-Navarra, S., Jensen, J., Schenk, F., Wahl, T., & Weisse, R. (2014). North Sea storminess from a novel storm surge record since AD 1843. *Journal of Climate*, 27(2014), 3582–3595. <https://doi.org/10.1175/JCLI-D-13-00427.1>
- Feng, J., & Jiang, W. (2015). Extreme water level analysis at three stations on the coast of the Northwestern Pacific Ocean. *Ocean Dynamics*, 65(11), 1383–1397. <https://doi.org/10.1007/s10236-015-0881-3>
- Feng, J., Von Storch, H., Jiang, W., & Weisse, R. (2015). Assessing changes in extreme sea levels along the coast of China. *Journal of Geophysical Research: Oceans*, 120, 8039–8051. <https://doi.org/10.1002/2015JC011336>
- Feng, X., & Tsimplis, M. N. (2014). Sea level extremes at the coasts of China. *Journal of Geophysical Research: Oceans*, 119, 1593–1608. <https://doi.org/10.1002/2013JC009607>
- Fritz, H. M., Blount, C. D., Thwin, S., Thu, M. K., & Chan, N. (2009). Cyclone Nargis storm surge in Myanmar. *Nature Geoscience*, 2(7), 448–449. <https://doi.org/10.1038/ngeo558>
- Guo, Y., Zhang, J., Zhang, L., & Shen, Y. (2009). Computational investigation of typhoon-induced storm surge in Hangzhou Bay, China. *Estuarine, Coastal and Shelf Science*, 85(4), 530–536. <https://doi.org/10.1016/j.ecss.2009.09.021>
- Ishii, M., Kimoto, M., Sakamoto, K., & Iwasaki, S.-I. (2006). Steric sea level changes estimated from historical ocean subsurface temperature and salinity analyses. *Journal of Oceanography*, 62(170), 1–16.
- Kernkamp, H. W. J., Dam, A. V., Stelling, G. S., de Goede, E. D., Van Dam, A., Stelling, G. S., & de Goede, E. D. (2011). Efficient scheme for the shallow water equations on unstructured grids with application to the Continental Shelf. *Ocean Dynamics*, 61(8), 1175–1188. <https://doi.org/10.1007/s10236-011-0423-6>
- Krueger, O., Schenk, F., Feser, F., & Weisse, R. (2013). Inconsistencies between long-term trends in storminess derived from the 20CR reanalysis and observations. *Journal of Climate*, 26(3), 868–874. <https://doi.org/10.1175/JCLI-D-12-00309.1>
- Marcos, M., Calafat, F. M., Berihuete, Á., & Dangendorf, S. (2015). Long-term variations in global sea level extremes. *Journal of Geophysical Research: Oceans*, 120, 8115–8134. <https://doi.org/10.1002/2015JC011173>
- Mellor, G. L. (2004). Users guide for ocean model. *Ocean Modelling*, 8544(June), 0710.
- Menéndez, M., & Woodworth, P. L. (2010). Changes in extreme high water levels based on a quasi-global tide-gauge data set. *Journal of Geophysical Research: Oceans*, 115, C10011. <https://doi.org/10.1029/2009JC005997>
- Merrifield, M. A., Genz, A. S., Kontoes, C. P., & Marra, J. J. (2013). Annual maximum water levels from tide gauges: Contributing factors and geographic patterns. *Journal of Geophysical Research: Oceans*, 118, 2535–2546. <https://doi.org/10.1002/jgrc.20173>
- Mori, N., Kato, M., Kim, S., Mase, H., Shibutani, Y., Takemi, T., . . . Yasuda, T. (2014). Local amplification of storm surge by Super Typhoon Haiyan in Leyte Gulf. *Geophysical Research Letters*, 41, 5106–5113. <https://doi.org/10.1002/2014GL060689>
- Muis, S., Verlaan, M., Winsemius, H. C., Aerts, J. C. J. H., & Ward, P. J. (2016). A global reanalysis of storm surges and extreme sea levels. *Nature Communications*, 7, 11969. <https://doi.org/10.1038/ncomms11969>
- Müller-Navarra, S. H., & Giese, H. (1999). Improvements of an empirical model to forecast wind surge in the German Bight. *Deutsche Hydrographische Zeitschrift*, 51(4), 385–405. <https://doi.org/10.1007/BF02764162>
- Pawlowicz, R., Beardsley, B., & Lentz, S. (2002). Classical tidal harmonic analysis including error estimates in MATLAB using TDE. *Computers and Geosciences*, 28(8), 929–937. [https://doi.org/10.1016/S0098-3004\(02\)00013-4](https://doi.org/10.1016/S0098-3004(02)00013-4)
- Pezeshki, S. R., Delaune, R. D., & Patrick, W. H. (1990). Flooding and saltwater intrusion: Potential effects on survival and productivity of wetland forests along the U.S. Gulf Coast. *Forest Ecology and Management*, 33–34(C), 287–301. [https://doi.org/10.1016/0378-1127\(90\)90199-L](https://doi.org/10.1016/0378-1127(90)90199-L)
- Pugh, D., & Woodworth, P. (2014). *Sea-level science understanding tides, surges, tsunamis and mean sea-level changes* (2nd ed., 407 p.). Cambridge, UK: Cambridge University Press.
- Rueda, A., Camus, P., Méndez, F. J., Tomás, A., & Luceño, A. (2016). An extreme value model for maximum wave heights based on weather types. *Journal of Geophysical Research: Oceans*, 121, 1262–1273. <https://doi.org/10.1002/2015JC010952>
- Wahl, T., & Chambers, D. P. (2016). Climate controls multidecadal variability in U.S. extreme sea level records. *Journal of Geophysical Research: Oceans*, 121, 1274–1290. <https://doi.org/10.1002/2015JC011057>
- Wang, X. L., Feng, Y., & Swail, V. R. (2012). North Atlantic wave height trends as reconstructed from the 20th century reanalysis. *Geophysical Research Letters*, 39, L18705. <https://doi.org/10.1029/2012GL053381>
- Wang, Y.-H., Lee, I.-H., & Wang, D.-P. (2005). Typhoon induced extreme coastal surge: A case study at Northeast Taiwan in 1994. *Journal of Coastal Research*, 213, 548–552. <https://doi.org/10.2112/03-0026.1>
- Woodworth, P. L., Hunter, J. R., Marcos, M., Caldwell, P., Menéndez, M., & Haigh, I. (2017). Towards a global higher-frequency sea level dataset. *Geoscience Data Journal*, 3, 50–59. <https://doi.org/10.1002/gdj3.42>
- Zhang, H., & Sheng, J. (2015). Examination of extreme sea levels due to storm surges and tides over the northwest Pacific Ocean. *Continental Shelf Research*, 93, 81–97. <https://doi.org/10.1016/j.csr.2014.12.001>
- Zhang, W.-Z., Hong, H.-S., Shang, S.-P., Chen, D.-W., & Chai, F. (2007). A two-way nested coupled tide-surge model for the Taiwan Strait. *Continental Shelf Research*, 27(10), 1548–1567. <https://doi.org/10.1016/j.csr.2007.01.018>